Deep Learning Enabled Fine-Grained Path Planning for Connected Vehicular Networks

Hanlin Wu, *Student Member, IEEE*, Haibo Zhou, *Senior Member, IEEE*, Jiwei Zhao, *Student Member, IEEE*, Yunting Xu, *Student Member, IEEE*, Bo Qian, *Student Member, IEEE*, and Xuemin Shen, *Fellow, IEEE*

***Abstract*—In this paper, to alleviate the ever-increasing traffic congestion in urban areas by accommodating higher road traffic, we develop traffic prediction framework together with path planning method for connected vehicular networks. First, through the employment of convolutional neural network (CNN) and residual unit (RN), deep learning (DL) based fine-grained traffic prediction algorithm is designed to obtain the spatial-temporal characteristics of vehicular traffic. The regionally fine-grained traffic prediction framework can realize real-time traffic prediction of future changing trends at each road with a high accuracy and reliability. Second, we propose a gridded path planning method by making use of the traffic prediction information. The accuracy of selected path, complexity of path calculation, and adaptive path adjustment are jointly taken into consideration by achieving the refined traffic regulation in different gridded section. Finally, we utilize the actual vehicle data from the city of Beijing and digital map on OpenStreetMap to validate the effectiveness and reliability of the proposed traffic prediction framework and path planning method. Simulation results demonstrate that the proposed approach is capable of relieving urban traffic congestion based on the existing roadway systems, which can provide methodological guidance for data-intensive traffic management.**

***Index Terms*—Connected vehicular networks, deep learning, path planning, spatio-temporal correlation, traffic prediction.**

I. Introduction

With the continuously increasing amount of private vehicles in urban areas, alleviating traffic congestion has become one of the most challenging issues for providing efficient traffic management and convenient travel experience [1], [2]. According to the statement of Urban Mobility Scorecard, traffic congestion in the United States causes people to spent nearly 6.8 billion hours more on travelling each year, aligned with a total economic cost estimated at 160 billion [3], [4]. Traditional attempt for relieving congestion situation is to expand traffic capacity by primarily broadening road space or building more transport infrastructures. However, due to the limited space in metropolitan city and the surging traffic that can easily saturate the road capacity, long-standing congestion continues to be the bane of many urban areas for decades [5]. As wireless communication significantly develops in recent years, cooperative traffic management solution via vehicle-to-everything (V2X) technologies open up a new window of opportunity to tame congestion based on the current layout of transportation system [6]–[8]. Besides, the advent of connected vehicles (CVs) equipped with intelligent cameras and sensors further enriches the way to obtain real-time traffic data, which makes the information intensive traffic regulation high on the list of possibilities [9]–[11].

Although the advancement of wireless communication and vehicular technologies is capable of collecting and transmitting sufficient traffic information, how to deal with sophisticated data processing problem under individualized scheduling task is still a matter of great difficulty [12]–[14]. With recent artificial intelligent (AI) technologies applied into the filed of intelligent transportation system (ITS), deep learning (DL) based traffic prediction has attracted widespread attention from both academia and industry [15]. On the one hand, urban traffic possesses salient characteristics such as local coherence and flow periodicity, which means that the urban traffic can be predicted beforehand by learning the spatio-temporal correlations from historical traffic observation [16]. On the other hand, classical artificial neural network in DL, like convolution neural network (CNN), can automatically and hierarchically capture the structural information of traffic flow through convolution operations [17]. Therefore, when it comes to the multi-dimensional data processing and computing for traffic prediction task, various neural networks can be leveraged to attain the changing trends of future traffic, with a high level of accuracy and reliability [18].

After obtaining the traffic prediction information, effective exploitation of the information can provide powerful assistance for creative data-driven scheduling strategies through taking current traffic situations and future changing trends into consideration [19]. Among all the potential solutions of relieving traffic congestion, the path planning, acting as a kind of effective method to prevent selfish driving choices which can worsen the inefficient traffic network, has gained lots of popularity for many classic algorithms such as Dijkstra algorithm [20] and

Manuscript received 1 August 2021; revised 4 February 2022 and 11 May 2022; accepted 23 May 2022. Date of publication 22 June 2022; date of current version 17 October 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 61871221, in part by the Innovation and Entrepreneurship of Jiangsu Province High-Level Talent Program, in part by the Summit of the Six Top Talents Program of Jiangsu Province, in part by the Program A for Outstanding Ph.D. Candidate of Nanjing University, and in part by the Natural Science and Engineering Research Council of Canada (NSERC). The review of this article was coordinated by Dr. Zhipeng Cai. *(Corresponding author: Haibo Zhou.)*

Hanlin Wu, Haibo Zhou, Jiwei Zhao, Yunting Xu, and Bo Qian are with the School of Electronic Science and Engineering, Nanjing University, Nanjing 210023, China (e-mail: hanlin_wu@smail.nju.edu.cn; haibozhou@nju.edu.cn; jw_zhao@smail.nju.edu.cn; yuntingxu@smail.nju.edu.cn; boqian@smail.nju.edu.cn).

Xuemin Shen is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: sshen@uwaterloo.ca).

Digital Object Identifier 10.1109/TVT.2022.3185249

A* algorithm [21]. Based on the acquired traffic information, path planning algorithm plays an a decisive role in the performance of traffic management efficiency [22]. However, most of the existing path planning methods are usually accompanied with remarkable computation complexity, especially under the city-scale path planning situation. Moreover, due to the dynamic changing environment brought by high-speed mobility of vehicles, it is difficult for them to satisfy the requirement of real-time adjustment so as to get with people's intention to travel as well as their associated travel behavior.

With the above-mentioned considerations, we first present a DL based fine-grained traffic prediction method to obtain the future traffic information of each urban road. After that, to improve the effectiveness of path planning and ensure low computation complexity, an innovative gridded path planning algorithm based on predicted traffic information is proposed to realize the refined vehicle scheduling in different gridded area. The main contributions of this paper are summarized as follows:

- We propose a fine-grained grid-level traffic prediction framework to realize the real-time traffic prediction of each road with a high level of accuracy and reliability, through which can significantly reduce the complexity of traffic prediction in large-scale road networks and provide effective guidance on vehicular scheduling.
- We introduce a gridded path planning method through jointly considering the accuracy of selected path, complexity of path calculation and real-time path adjustment, which can provide general and effective guidance on realtime vehicular path planning for some suddenly emerged traffic scenarios, such as traffic emergencies, changeable weather, frequent road construction, etc.
- We verify the proposed fine-grained traffic prediction framework and gridded path planning method using the actual vehicle data in practical traffic system and digital map from OpenStreetMap. Simulation results confirm that our solution can promote efficient usage of roadway systems and better deal with congested problem.

The rest of this paper is organized as follows. Section II introduces the related work of traffic prediction and path planning. The system model is discussed in Section III. Section IV presents the proposed fine-grained traffic prediction framework and gridded path planning is described in Section V. Section VI demonstrates the simulation results. Finally, this paper is concluded in Section VII.

II. Related Works

A. Traffic Prediction

Traffic prediction has been receiving remarkable research attentions in the field of ITS over the past decades [23]. With recent employment of superior AI technologies, traffic prediction has shown great potentials in providing regulation assisted applications such as vehicle navigation, congestion management and traffic equilibrium, etc [24]. The state-of-the-arts AI enabled traffic prediction methods can be broadly classified into two categories, i.e., model-driven and data-driven methods. In the context of model-driven prediction methods, traffic models of road network are computationally established so as to analyzes the performance of scheduling solution through simulation validation. Ko *et al.* [25] and Bolshinsky *et al.* [26] introduced the comprehensive model-driven traffic prediction methods combined with DL algorithms such as regression model and Markov chain model in detail. Abadi *et al.* [27] proposed an autoregressive model to predict the traffic flow by using a dynamic traffic simulator to generate the traffic flow data. As the ever-increasing number of vehicle becomes more and more common on the road, the processing of sophisticated traffic data requires enormous device computing capability for model-driven methods. Therefore, data-driven traffic prediction emerges as an efficient solution, especially in dynamic changing environment. Ma *et al.* [28] applied CNN to traffic prediction by taking traffic images as network input to follow abstract traffic feature extraction and conduct network-wide traffic speed prediction. Chen *et al.* [16] presented a different use of CNN to solve the traffic prediction problem. The traffic data were folded to form a two-dimensional matrix considering the traffic in the past and similar historical patterns simultaneously.

However, these traffic prediction methods mentioned above only focus on predicting future traffic from the spatial perspective or the temporal perspective, without jointly considering their inner correlations. In order to capture the diversified characteristics of urban traffic, Zhang *et al.* [29] proposed a deep spatio-temporal residual network composed of convolution and residual unit. The convolution operation was applied to capture the spatial near dependencies, and residual unit was used to treat information in adjacent time intervals as multiple channels for capture the temporal dependency as well as ensure the training effectiveness. To further improve the accuracy of prediction, Guo *et al.* [30] proposed a 3D network (ST-3DNet) exploiting 3-dimensional convolution to avoid the deficiency of temporal information. Since traffic prediction can provide reliable traffic guidance information and help to improve the safety and efficiency of ITS, Li *et al.* [18] analyzed the inflow and outflow traffic of different areas based on traffic prediction, which can balance traffic load and finally promote the application of intelligent traffic management.

B. Path Planning

In order to monitor and manage the dynamic traffic conditions actively, path planning acts as an innovative way to influence people's need to travel as well as their associated driving behavior, with the purpose of promoting efficient use of existing roadway systems and better handling of the vehicle congestion [31]. Among the classical path planning algorithms, Dijkstra algorithm finds the shortest distance from departure point to destination by iterating node-selecting computation in the search space [20], while the A* algorithm reduces the calculation time of the road selection through always solving the core problem of the path planning algorithm [21].

In order to decrease the computation complexity of large-scale path planning, some hierarchical path planning algorithms divided the road network into different regions and levels, for reducing the data volume of the road network and the search space of each path selection size [32]. Mainali *et al.* [33] proposed a dynamical path planning algorithm based on the traffic

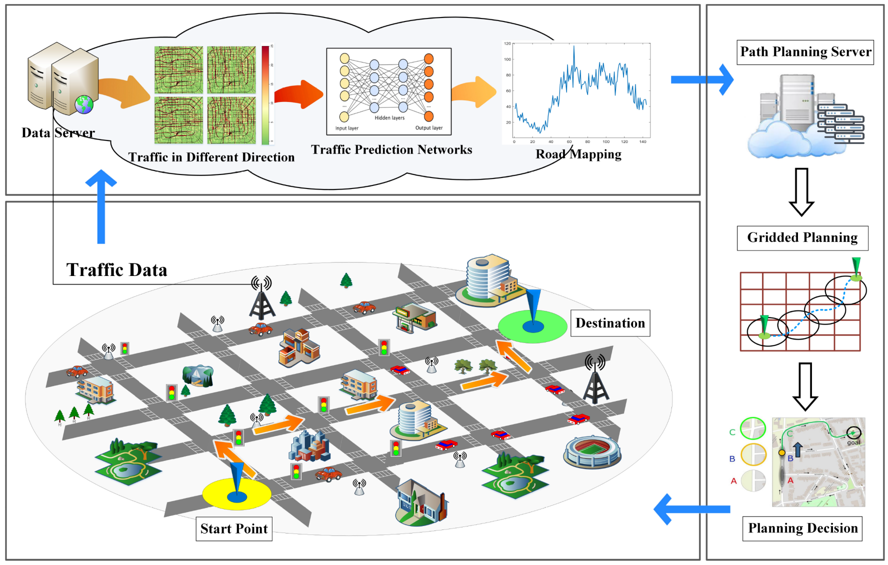


Fig. 1. The scenario of gridded path planning services based on traffic prediction for connected vehicular networks.

conditions and used a recursive method to solve the problem of vehicle travelling. Wang *et al.* [34] and Guo *et al.* [35] leveraged adjustable path pattern to deal with traffic congestion caused by some emergencies in urban traffic. However, although current path planning methods can alleviate traffic congestion to a certain degree, they cannot adapt to the rapidly changing traffic conditions while satisfy the computation need for constantly modifying the navigation path of each vehicle. Therefore, in this paper, a gridded path planning method is presented to effectively solve the congestion problem in the scenario of urban road network. With the traffic prediction information considered in path planning, refined vehicle scheduling in different gridded section is able to reduce computation overhead, make real-time regulation and realize efficient traffic management.

III. System Model and Preliminaries

In this section, we formulate the DL based fine-grained traffic prediction problem and outline the model of gridded path planning for global connected urban road system.

A. Overview

Fig. 1 illustrates the scenes of gridded path planning services for vehicles based on traffic prediction to help vehicles find the shortest route to their destination. The process consists of the following three stages:

1) *Traffic Data Perception and Processing:* First of all, with the advanced vehicle-to-everything (V2X) technologies, a large number of sensing devices existing in connected vehicular networks, such as GPS on vehicles, cameras placed on the edge of roads, participate in the perception of urban traffic and collect traffic data. Then, the collected traffic data are transmitted to the cloud server for traffic prediction.
2) *Deep Learning Based Fine-grained Traffic Prediction:* Secondly, after obtaining historical traffic observations, deep learning is used to learn the temporal and spatial correlations in urban traffic. As a consequence, the traffic situation for a certain time in the future can be predicted, from which we can receive the information for vehicles' path planning.
3) *Gridded Path Planning for Vehicles:* Finally, predicted traffic information from deep learning based traffic prediction can be considered in vehicles's path planning. Due to high complexity of city-level path planning and the tradeoff of traffic prediction between accuracy and multi-step ahead prediction, gridded path planning plans the path for the vehicle of the area where the vehicle is currently located.

Based on the above description, we first difine the urban road network as $G = \{R, Q\}$, in which R is the directed graph of the road and Q is the intersection set of urban traffic. For example, R_{ij} represents the road section from intersection Q_i to Q_j. In addition, for each road section R_{ij}, we use L_{ij} to represent its length. It is worth noting that, road section has directional characteristics which means $R_{ij} \neq R_{ji}$, while their length are the same, so $L_{ij} = L_{ji}$. Besides, in urban road network, traffic management center periodically collects the real-time information of vehicles. The real-time information of each vehicle can be expressed by $\boldsymbol{p_i} = \{t_i, lon_i, lat_i, \vec{v_i}\}$, where p_i means the related parameters of vehicle i driving on the road network, t_i is the time when the vehicle uploads its information, lon_i and lat_i is the latitude and longitude of vehicle i, and $\vec{v_i}$ is the speed vector of the vehicle which contains driving speed and the direction. At last, traffic management center unifies the collected information to form the sets of vehicle information at different time intervals, which can be expressed by

$$\boldsymbol{P}_t = \left\{ p_k | t - \frac{t_s}{2} \leq t_k \leq t + \frac{t_s}{2} \right\},$$

where $\boldsymbol{P}_t$ means the collected information of all vehicles driving on road network at time interval $(t - \frac{t_s}{2}, t + \frac{t_s}{2})$, p_k represents

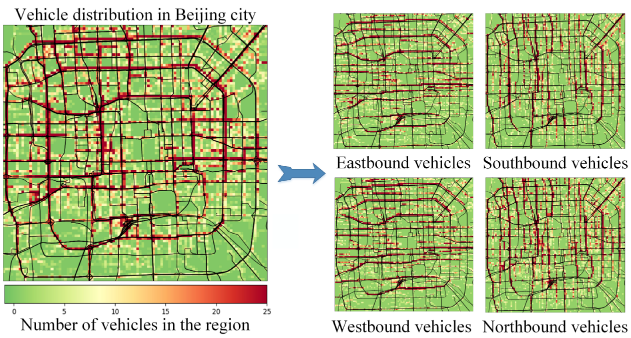


Fig. 2. The vehicle data is distinguished according to the driving direction to form a tensor F_t.

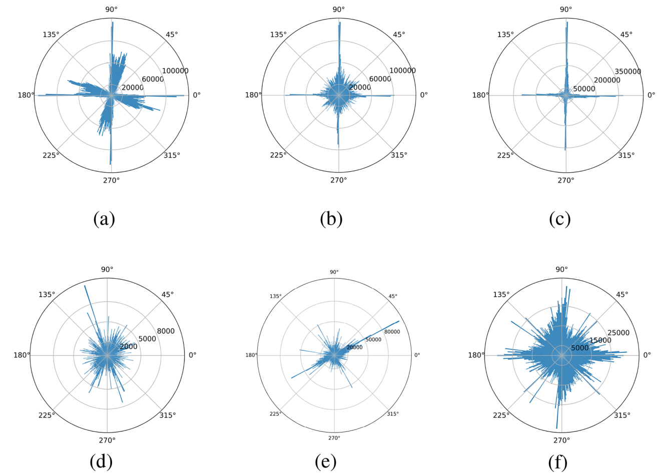


Fig. 3. The analysis of road directions of some classic cities in China and other countries.

a certain vehicle k driving on the road network, t_k is the time when the vehicle k uploads its information at a certain time, and t_s represents the value of the time interval.

B. Formulation of Traffic Prediction

In practical traffic system, the locations of vehicles are constrained by the topological relationship of urban roads, so the traffic flow exhibits a high degree of temporal and spatial correlation, which means that it's highly predictable for the traffic flow. In this paper, we study the problem of traffic prediction on Beijing's road network. As shown in Fig. 2, we first define the square region of the urban road network and directional traffic flow, and then formulate the traffic prediction problem.

Definition 1 (Square Region): We divide the entire urban traffic area into $M * N$ square regions according to the latitude and longitude, and each square region represents the traffic flow in this area. For instance, $x_{m,n}$ represents the traffic flow in the square region of the m^{th} row and n^{th} column.

Definition 2 (Direction of Traffic Flow): Considering that the driving direction is an important condition for determing which road a vehicle belongs to, we divide the directions of vehicles into different sectionss. We use $S = \{s_1, s_2, s_3 \ldots\}$ to express the sections of different driving directions, and traffic flow in a certain square region will be divided into different sections, which can be expressed by $x_{m,n} = \sum x_{m,n}^{s_i}$.

Considering the distribution of the Beijing road network shown in Fig. 2, we divide the traffic flow in each square region into four directions of south, east, north and west, according to their driving directions. Therefore, $S = \{s_1, s_2, s_3, s_4\}$. In detail, the driving direction of vehicle is expressed by azimuth, the angle between driving direction of vehicles and the north in a clockwise way. As a consequence, in this study $\{s_1, s_2, s_3, s_4\}$ are $(45, 135), (135, 225), (225, 315), (0, 45) \cup (315, 360)$.

To further illustrate the rationality and universality of Definition 2 due to the fact that the road network of Beijing is in a typical square shape while other cities may not be the same, we analyze the direction distribution of the roads in different cities in Fig. 3. The horizontal and vertical axes of Fig. 3 respectively represent the azimuth range of the roads and the amount of the roads. Then we can find that the direction distribution of other cities has a strong regularity although their road networks are not as square as Beijing. Therefore, based on the direction distribution, we can regard the road directions with very few roads as boundaries, and divide the roads into appropriate direction sections.

After processing the vehicle information according to the Definitions 1 and 2, we can get a directional square maps of the entire urban road network. For example, the traffic flow at time intervals t in a certain driving direction s_i at the location of the m^{th} rows and the n^{th} columns in the square maps can be expressed by $x_{m,n}^{s_i,t}$, which is definned as follow:

$$x_{m,n}^{s_i,t} = \left| \{p_k | \vec{v_k} \in s_k \wedge p_k \in x_{m,n}\} \right| \quad s_k \in S\,,\, p_k \in \boldsymbol{P}_t, \tag{1}$$

where $x_{m,n}^{s_i,t}$ means the number of vehicles in region (m, n) with driving direction in s_i at time interval t. $|.|$ represents the element amount of the collection, and p_k is the collected information of vehicle k, which has been descripted above. $p_k \in x_{m,n}$ denotes that vehicle k is within the square region $x_{m,n}$, $\vec{v_k} \in s_i$ means that the driving direction of vehicle k is within s_i, and $p_k \in \boldsymbol{P}_t$ means that this vehicle information belongs to the information set $\boldsymbol{P}_t$ at time interval t.

Then we can get a square map of the urban road network with 4 directions at each time interval, which can be summarized as a tensor $X \in \mathbb{R}^{4*M*N}$. Based on the above definitions and descriptions, the problem of predicting the traffic using the historical traffic information can be expressed as below:

Given

$$\{X_t | t = t_0, t_0 - t_s, t_0 - 2t_s, \cdots\},$$

Predict

$$\{X_t | t = t_0 + t_s, t_0 + 2t_s, \cdots\}.$$

C. Formulation of Path Planning

After traffic prediction, in this section, we study the problem of path planning for vehicles, in which traffic information plays a significant role. So we define three kinds of traffic information for path planning of vehicles according to their acquisition time as shown below.

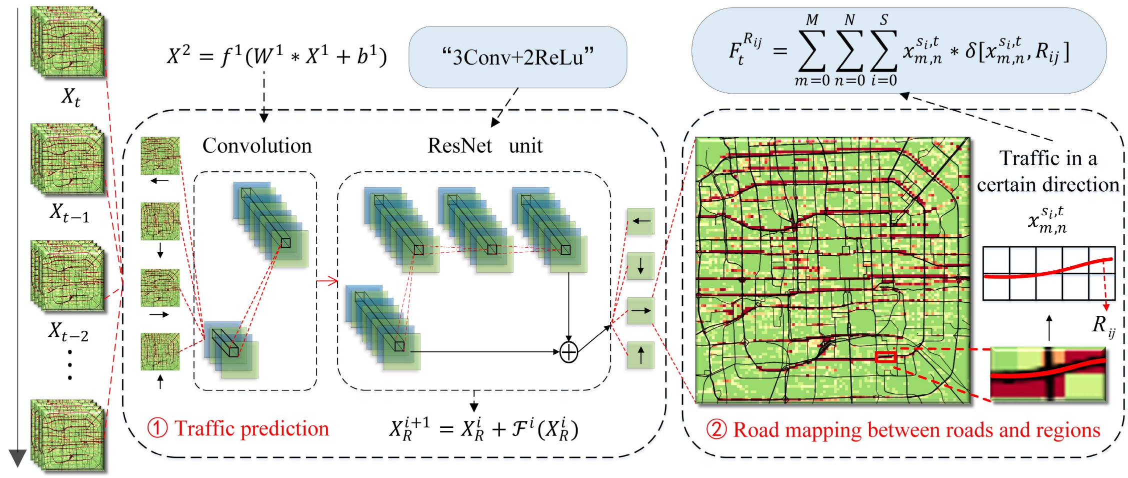


Fig. 4. Neural network architecture and road mapping relationship for regionally fine-grained traffic prediction.

Definition 3 (Historical Information): Historical Information represents the traffic information collected a long time ago, such as the traffic information collected in the last week. The urban traffic is periodic due to its spatio-temporal correlation, which means that historical information has a certain guiding significance at the path planning of vehicles.

Definition 4 (Real-time Information): Real-time Information is the information that is collected currently. In fact, because of the time delay during information collection, processing and transmission, the latest traffic information available to vehicles may be a few minutes ago. Therefore, the real-time traffic information commonly referred to is short-term historical information.

Definition 5 (Prediction Information): Prediction Information means the traffic information for a certain time in the future. By knowing the prediction information, we understand the changing trends and future conditions of urban traffic. However, prediction information also has the accuracy problem need to be considered.

After defining the traffic information of vehicles, we study the problem of vehicles traveling in urban road network. There are many factors to consider when choosing a path for a vehicle traveling in urban road network, and we consider the time-dependent road network in this study which means that the purpose of the vehicle's path selection is to minimize the time consuming from origin to destination. Therefore, the cost of a vehicle passing through a certain road section R_{ij} is the time it takes, which can be expressed by $T(i,j)$. Formally, the formulation of the problem of a vehicle traveling from origin to destination can be described as follows:

$$\begin{aligned} \min \quad & C = \sum_{R_{ij} \in f^{od}} T(i,j) \\ s.t. \quad & R_{ij} \in R \\ & f^{od} \geq 0, T(i,j) \geq 0, \end{aligned} \tag{2}$$

where o, d represents the origin and the destination, f^{od} is a certain path from origin to destination, C represents the total time consuming for vehicle passing through the path f^{od}.

IV. Traffic Prediction and Speed Estimation

In this section, we propose a directional-spatio-temporal residual neural networks based fine-grained traffic prediction method, as shown in Fig. 4, which includes traffic prediction using CNN based residual networks and the evaluation of average driving speed of each road.

A. Traffic Prediction for Fine-Grained Directional Regions

Because of the temporal and spatial correlations of the traffic flow, deep learning based neural networks can be effectively used in traffic prediction by learning the historical traffic observation. The specific structure of the neural network used here is shown in Fig. 4 and we call it Direction-Spatio-Temporal ResNet (DST-ResNet). The main structures of DST-ResNet includes CNN and residual unit (RU).

Although the entire city covers a very large area, vehicle's trajectory is constrained by factors such as road topology and speed restriction. Therefore, the locations of vehicles are related in adjacent time, meaning that the traffic flow is highly predictable [16]. Besides, as mentioned before, CNN has the advantages of local feature extraction and weight sharing. In CNN, convolution kernel is the basic composition, through which each output only connects to local values and this local connectivity can help to the local feature extraction. Moreover, CNN can reduce the training complexity of neural networks by its sharing weights. Thereby, CNN is utilized in DST-ResNet to learn the spatial correlation in the traffic flow. The operation of a layer of CNN is represented as the following formula:

$$X^{(l+1)} = f^{(l)}\left(W^{(l)} * X^{(l)} + b^{(l)}\right), \tag{3}$$

where $*$ means the operation of convolution, $W^{(l)}$ and $b^{(l)}$ are the learnable parameters of layer l in the neural network. $(W^{(l)} * X^{(l)} + b^{(l)})$ and $f^{(l)}$ are respectively the linear and avtivation function, which is a non-linear operation. In DST-ResNet, we set $f^{(l)}$ as the ReLu function, which can be expressed by $f^{(l)}(\theta) = max(0, \theta)$.

So as to learn the spatial correlation in traffic flow amog multiple regions and improve the traffic prediction accuracy, a multi-layered CNN is necessary in this study. Owing to the limited size of convolution kernel, only the spatial correlation among the surrounding regions can be learned in each layer of CNN. Consequently, a multi-layered CNN is conducive to learning the spatial correlation of extensive traffic flow. However, a deeper neural network will bring harmful effects on the training effect of the neural network and easily lead to an explosion gradient, disappear and other issues. With the purpose of solving the problem may caused by deeper neural networks, the residual unit, which has been proven to achieve great success in training ultra-deep neural networks with more than 1000 layers [36], is used in our study. Because of the connection of the input and output separated by multiple layers of neural network by residual unit, a deeper neural network does not lead to a huge increase or decrease in errors. The structure of the residual unit can be expressed by the following formula:

$$X_R^{(l+1)} = X_R^{(l)} + \mathcal{F}\left(X_R^{(l)}\right), \tag{4}$$

where $X_R^{(l)}$ and $X_R^{(l+1)}$ are the input and output of the l^{th} residual unit. $\mathcal{F}$ is a specific structure in residual unit consists of three "ReLu + Convolution".

Based on the structure description of DST-ResNet above, we use historical traffic observations to predict future traffic. Firstly, we convert the collected vehicle information into 4-channel images at different time intervals in line with Definitions 1 and 2. After that, in order to learn the temporal correlation of traffic flow among adjacent time intervals, we utilize the image of traffic flow at three historical adjacent time intervals at the same time as the intput of DST-ResNet. Finally, the input of these three images passing through DST-ResNet are fused into the final prediction result for the next time interval.

Besides, during the training process, a loss function is indispensable. In this study, we utilize the mean square error between the prediction result x_t^* and actual vehicle data x_t as the loss function, which can be expressed as follows:

$$Loss = \sum_{m=0}^{M}\sum_{n=0}^{N}\sum_{i=0}^{S} \frac{|x_{m,n}^{s_i,t} - \check{x}_{m,n}^{s_i,t}|^2}{M * N * S}, \tag{5}$$

where M and N are the size of grid according to Definition 1. S represents the amount of direction sections according to the Definition 2. As mentioned before, the driving direction of vehicles is divided into four sections $\{s_0, s_1, s_2, s_3\}$, so $S = 3$.

B. Evaluation of Average Vehicle Speed

After the traffic prediction for fine-grained directional regions descriibed above, in this section, we study how to evaluate the traffic flow of each single road based on the predicted traffic flow in different regions with different directions, which is similar to road matching. In traditional road matching, each individual vehicle is usually matched with the roads based on their location and direction, which will lead to a large amount of calculations and can not meet the real-time requirements. In this study, we match the directional square regions with roads and evaluate the average driving speed of each single road as shown in Fig. 4.

First of all, we establish a mapping relationship between the directional square regions and the roads according to the distance between them and the consistency of their directions. In this study, the distance between a square region and a certain road represent the distance from the center of the directional square region to the road. And consistency of their directions is determined by the direction section that the road belongs to. As a consequence, the distance between the square region (m, n) and the road R_{ij} can be expressed by the following formula:

$$d_{R_{ij}}^{(m,n)} = \begin{cases} d_{Q_i}^{m,n} & \left(d_{Q_i}^{m,n}\right)^2 + (L_{ij})^2 < \left(d_{Q_j}^{m,n}\right)^2 \\ d_{Q_j}^{m,n} & \left(d_{Q_j}^{m,n}\right)^2 + (L_{ij})^2 > \left(d_{Q_i}^{m,n}\right)^2 \\ H(\theta) & else \end{cases}, \tag{6}$$

where $d_{R_{ij}}^{(m,n)}$ represents the distance from regioin (m, n) to road R_{ij}, $d_{Q_i}^{m,n}$ and $d_{Q_j}^{m,n}$ means the distance from regioin (m, n) to the intersection Q_i, Q_j, and $H(\theta)$ is Helen's formula, which is

$$\begin{cases} H(\theta) = \frac{\sqrt{\theta\left(\theta - d_{Q_i}^{m,n}\right)\left(\theta - d_{Q_j}^{m,n}\right)(\theta - L_{ij})}}{L_{ij}} \\ \theta = \frac{\left(d_{Q_i}^{m,n} + d_{Q_j}^{m,n} + L_{ij}\right)}{2} \end{cases}. \tag{7}$$

Then, the mapping relationship between the directional square region (m, n) and the road R_{ij} can be expressed as follows:

$$x_{m,n}^{s_i,t} \in R_{ij} \quad if \quad \left(d_{R_{ij}}^{(m,n)} < \eta \wedge s_{R_{ij}} \in s_i\right), \tag{8}$$

where η is the threshold of the distance between the square region and the road, $s_{R_{ij}}$ is the direction of the road R_{ij}, $x_{m,n}^{s_i,t} \in R_{ij}$ means there is a mapping relationship between square region (m, n) and road R_{ij}. $d_{R_{ij}}^{(m,n)} < \eta$ means that the distance between square region (m, n) and road R_{ij} is smaller than a given threshold and $s_{R_{ij}} \in s_i$ represents that the direction of road R_{ij} matches the direction of square region (m, n).

Based on the above description and Definitions 1, 2, we can evaluate the traffic flow of each road according to the matching relationship. Obviously, the accuracy of evaluation is closely related to the size of a square region. Firstly, the size of a square region should be small enough to meet the evaluation accuracy of most roads, otherwise a square region in one direction may have a matching relationship with multiple roads resulting in the inaccuracy of evaluation. Besides, when the size of a square region tends to be zero, there is at most one vehicle in each square region and the matching between roads and regions becomes the matching between roads and every single vehicle, as shown in the following formula:

$$x_{m,n}^{s_i,t} \in \{0, 1\} \quad M, N \to \infty. \tag{9}$$

However, the reduction of the size of a square region leads to the increased temporal and spatial complexity of traffic prediction. Moreover, when the size of a square region is too small, there will exist a lot of invalid regions, which will result in the damage of learning spatial correlation among regions. Therefore, considering both the accuracy of traffic prediction and evaluation of traffic flow, M, N should take appropriate values based on the coverage area of the city and the distance between different roads in the same direction. Specifically, we set $M = N = 100$ in this work after comparing other multiple values.

Finally, based on the established mapping relationship between directional square regions and the roads, the traffic flow of road R_{ij} at time t can be expressed by the following formula:

$$F_t^{R_{ij}} = \sum_{m=0}^{M} \sum_{n=0}^{N} \sum_{i=0}^{S} x_{m,n}^{s_i,t} * \delta\left[x_{m,n}^{s_i}, R_{ij}\right], \tag{10}$$

where $F_t^{R_{ij}}$ is the traffic flow of road R_{ij} at time interval t, δ represents the mapping relationship between the square region $x_{m,n}^{s_i,t}$ and road R_{ij}, as shown in the following formula:

$$\delta\left[x_{m,n}^{s_i}, R_{ij}\right] = \begin{cases} 1 & x_{m,n}^{s_i} \in R_{ij} \\ 0 & x_{m,n}^{s_i} \notin R_{ij} \end{cases}. \tag{11}$$

After the process descripted above, we can obtain the traffic flow of each single road at each time interval. Furthermore, the average speed for vehicles driving on a certain road is closely related to the traffic flow of the road [37]. The relationship between the traffic flow and average driving speed of a certain road is shown in the following formula:

$$v = v_{max}\left(1 - \lambda/\lambda_{max}\right) \tag{12}$$

where v_{max}, λ_{max} respectively represents the maximum speed and maximum capacity of the road, whose values are determined by the amount of lanes and speed limit of the road, and v, λ is the average driving speed and traffic flow of the road.

After that, the average driving speed for vehicles on each road in each time interval can be obtained, and we define the average speed calculated by historical information as V^h and the average speed calculated by prediction information as V^p. On the basis of the prediction of traffic flow and average driving speed, traffic management center can provide better path planning services for vehicles.

V. Path Planning Based on Traffic Prediction

In this section, we propose a gridded path planning method based on traffic prediction to select better path with less amount of calculation.

In the following, we first study adding traffic prediction to path calculation of path planning. Then, we propose a gridded path planning method, which not only use prediction information to find a better path but also reduce the amount of calculation.

A. *Traffic Prediction in Path Planning*

In traditional path planning, the time cost for vehicles to pass the road R_{ij} at time t_0 is usually calculated by $T_h(i,j,t_0) = \frac{L_{ij}}{V_{ij}^h(t_0)}$, in which $T_h(i,j,t_0)$ and $V_{ij}^h(t_0)$ means the time cost and average driving speed without traffic prediction. After adding traffic prediction into path planning, the time cost for vehicles to pass the road can be calculated as a dynamic process. Therefore, the time cost for vehicles to pass the road R_{ij} at time t_0 can be expressed by the following formula:

$$L_{ij} = \int_{t_0}^{t_0+T_p(i,j,t_0)} V_{ij}^p(t)dt, \tag{13}$$

where $T_p(i,j,t_0)$ and $V_{ij}^p(t)$ represents the time cost and the average travel speed for vehicles to pass the road R_{ij} at time t_0 adding traffic prediction.

Considering that the prediction information obtained by traffic prediction is discrete and the time intervals of traffic prediction are relatively short, we assume that the average speed for vehicles driving on a certain road varies linearly within a time interval. If a time interval is t_s, average travel speed of road R_{ij} at z^{th} time interval according to predicted information is $V_{ij}^p[z]$, then the average travel speed of the road R_{ij} at time t can be expressed by the following formula:

$$V_{ij}^p(t) = V_{ij}^p[z] + k * (t - z * t_s) \quad t < (z+1) * t_s, \tag{14}$$

where z is the first time interval less than t, and k represents the slope between the average travel speed $V_{ij}^p[z]$ and $V_{ij}^p[z+1]$, so

$$\begin{cases} k = \frac{V_{ij}^p[z+1] - V_{ij}^p[z]}{t_s} \\ z = \lfloor t/t_s \rfloor. \end{cases} \tag{15}$$

As a consequence, combining (13), (14), and (15), the time cost for vehicles passing a certain road R_{ij} at time t_0 can be calculated. For example, if $\lfloor (t_0 + T_p)/t_s \rfloor = \lfloor t_0/t_s \rfloor$, the time cost for vehicles passing a certain road R_{ij} at time t_0 can be expressed by

$$T_p(i,j,t_0) = \sqrt{\frac{2k_0 L_{ij} + V_{ij}^p(t_0)^2}{k_0^2}} - \frac{V_{ij}^p(t_0)}{k_0}, \tag{16}$$

where k_0 represensts the slope between the average travel speed $V^p[z_0]$ and $V^p[z_0+1]$.

When $\lfloor (t_0 + T_p)/t_s \rfloor > \lfloor t_0/t_s \rfloor$, the time cost for vehicles passing a certain road R_{ij} at time t_0 becomes the following formula:

$$\begin{cases} T_p(i,j,t_0) = \sqrt{\frac{2k_p\Delta + V_{ij}^p[z_p]^2}{k_p^2}} - \frac{V_{ij}^p[z_p]}{k_p} + z_p * t_s - t_0 \\ \Delta = L_{ij} - (z_1 * t_s - t_0)\frac{V^p[z_1] + V^p(t_0)}{2} \\ \qquad - \sum_{z=z_1}^{z_p-1} \frac{V^P[z] + V^P[z+1]}{2} \end{cases}, \tag{17}$$

where $z_p = \lfloor (t_0 + T_p)/t_s \rfloor$, z_1 means the next time interval of z_0 and k_p represents the slope between the average travel speed $V^p[z_p]$ and $V^p[z_p+1]$.

In this way, we apply traffic prediction to path planning and turn the calculation of time cost into a dynamic process, helping to calculate the best path more accurately. Then, we can calculate the shortest path from the origin to the destination through the global query method, such as the classic Dijkstra algorithm.

However, for the path planning in the entire city, the global query is unable to satisfy the real-time calculation of path planning using prediction information when destination is too far away. Besides, far destination means the need for multistep-ahead traffic prediction, which can lead to the decreased accuracy of traffic prediction. So in the next section, we design a gridded path planning method based on the traffic prediction

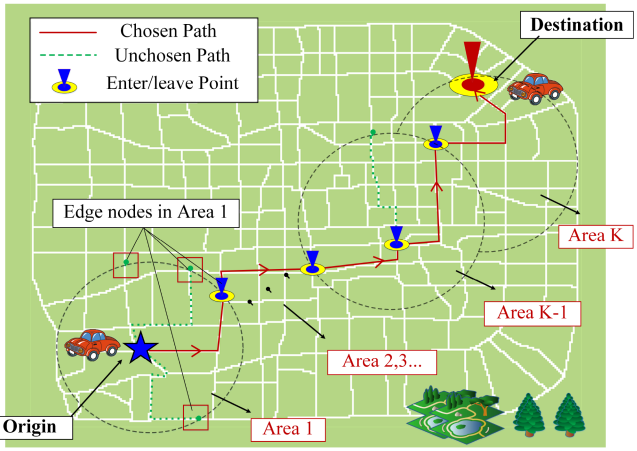


Fig. 5. The entire process of the gridded path planning algorithm.

Algorithm 1: Time Estimation to Destination.

Input: Road network topology matrix R, L, Q
Historical information V^h
Origin and destination o, d
Output: Estimated time to destination $\mathcal{T}_d$
Divide the searching area $\mathbb{B}$ according to o, d:
$\mathbb{B} : \{b_i | \, |o, b_i| + |b_i, d| \leq \mu_0 |o, d| \quad b_i \in Q\}$.
Calculate and save the estimated time $\mathcal{T}_d$ to destination for each node:
for each $j = 1, 2, 3 \cdots$ **do**
  Calculate the nodes contained in the l_j layer according to formula (22).
  for each $b_{i,l_j} \in l_j$: **do**
    Calculate $\mathcal{T}_d(b_{i,l_j})$ by formula (23).
  end for
end for

considering both the demand of real-time calculation and the accuracy of traffic prediction.

B. Gridded Path Planning

As shown in Fig. 5, the process of gridded path planning executes as follows. Firstly, a local area called grid area is divided according to the current location of the vehicle. Then, if the destination is within the grid area, the minimum time-consuming path is calculated according to the prediction information. Otherwise, only the path within the grid area is planned based on the shortest time from origin to the edge nodes of the grid area and the estimated time from the edge nodes to the destination. Finally, the grid area will be divided again and the path will be recalculated when vehicle is about to finish the path within the grid area until the destination is within the grid area.

Therefore, the whole gridded path planning method can be divided into three parts: division of the gridded area, travel time estimation, and path planning in gridded area.

1) Division of Gridded Area: In this study, we define a local area around where the vehicle is located called grid area and only the path in this area will be planned each time, which can ensure the low complexity to meet the needs of real-time computing and avoid the decreased accuracy of prediction information caused by multistep-ahead traffic prediction.

In detail, for a path planning request from the origin o to destination d at time t_0, we simply define the nodes that in the range of a circle with the center of the origin o and radius r as the grided area, which can be expressed by the following formula:

$$\mathbb{A}_{grid} : \{a_i| \quad |o, a_i| \leq r \quad o, a_i \in Q\} . \tag{18}$$

As for the size of the radius r, in order to make full use of prediction information, we match r with the time interval of traffic prediction, which can be expressed by the following formula:

$$r = v_{max} * t_s, \tag{19}$$

where v_{max} is the speed limit of vehicle's current area, and t_s represents the size of time interval of traffic prediction.

After the division of grided area, the time cost for vehicle to each node in the gridded area can be iteratively calculated using prediction information by referring to Dijkstra algorithm. When iterating to a certain node a_i in $\mathbb{A}_{grid}$, the iteration calculation can be expressed by the following formula:

$$\begin{cases} \mathcal{T}_o(a_i^{(j)}) = \min \left\{ \mathcal{T}_o(a_i^{(j)}), \mathcal{T}_o(a_i) + T_p(a_i, a_i^{(j)}, t_{a_i}) \right\} \\ t_{a_i} = t_0 + \mathcal{T}_o(a_i) \end{cases}, \tag{20}$$

where $a_i^{(j)}$ represents the neighbor node of a_i, $\mathcal{T}_o(a_i)$ is the time cost from origin to node a_i, and t_{a_i} means the time vehicle arrive at node a_i.

Then, if destination d is within the gridded area, we can get the precise path to the destination according to $\mathcal{T}_o(d), d \in \mathbb{A}_{grid}$.

2) Travel Time Estimation: However, for the grid area with limited scope, a destination farther away may not locate in this area. In this situation, only planning the path within the grid area can not guarantee the right path to destination. In this study, we propose a estimated time method to estimate the time cost from the edge node of grid area to destination and serve as a reference for path selection when destination is not within the grid area. Detailly, we use dynamic programming method to calculate the estimation time from the nodes to destination by historical traffic information.

First of all, we define the searching node set as $\mathbb{B} = \{b_1, b_2, \cdots b_i\}$. In detail, we divide a ellipse area based on the location of origin and destination, which is a classic path search area. Thus, we have

$$\mathbb{B} : \{b_i | \, |o, b_i| + |b_i, d| \leq \mu_0 |o, d| \quad b_i \in Q\} , \tag{21}$$

where o, d represents origin and destination of the vehicle travel, $|.|$ means a straight line distance, and μ_0 is a constant.

Secondly, we layer the node in $\mathbb{B}$ based on the number of nodes from it to the destination. For example, the neighbor node of the destination is the first layer and the destination is expressed by

l_0. The specific representation of each layer is as follows:

$$
\begin{aligned}
&l_1: \{b_i | b_i \in neighbors\ of\ d\} \\
&l_2: \{b_i | b_i \in neighbors\ of\ nodes\ in\ l_1\} - l_0 \\
&\qquad \vdots \quad \vdots \quad \vdots \\
&l_j: \{b_i | b_i \in neighbors\ of\ nodes\ in\ l_{j-1}\} - l_{j-2}. \qquad (22)
\end{aligned}
$$

Finally, we calculate the estimated time to the destination for each node in each layer iteratively starting from the first layer. If a node in the layer l_j is b_{0,l_j} and its neighbor nodes in the upper layer are defined as $b_{0,l_{j-1}}^{(i)} (i = 1, 2, 3 \ldots)$, then the estimated time from this node to the destination can be expressed as follows:

$$
\begin{aligned}
&\mathcal{T}_d(b_{0,l_j}) \\
&= \min \left[\mathcal{T}_d \left(b_{0,l_{j-1}}^{(i)} \right) + T_h \left(b_{0,l_j}, b_{0,l_{j-1}}^{(i)}, t_0' \right), i = 1, 2, 3 \ldots \right], \qquad (23)
\end{aligned}
$$

where $\mathcal{T}_d(b_{0,l_j})$ is used to represent the estimated time to the destination from b_{0,l_j}, and $T_h(b_{0,l_j}, b_{0,l_{j-1}}^{(i)}, t_0')$ means the calculation time from ndoe $b_{0,l_{j-1}}^{(i)}$ to b_{0,l_j} according to the historical traffic information. Specially, t_0' represents the time close to the traffic situation at t_0 in the historical traffic information, such as time t_0 at the same day last week.

So in this way, we can obtain the estimation time to destination from the nodes between origin and destination. Moreover, we use historical information to calculate this estimated time, thereby we can reduce the amount of real-time calculation through offline pre-calculation.

3) Path Planning in Gridded Area: Finally, the path in the Gridded area can be calculated accurately based on the traffic prediction. If the destination is within the gridded area, the path can be calculated easily using prediction information. Otherwise, the selection of the path need to consider both the accurate time from origin to the edge nodes of gridded area using prediction information and the estimation time from the edge nodes to destination.

Specifically, we define the edge nodes of the gridded area as $a_e^{(i)}$, so the path selection can be expressed by the following formula:

$$
Path^*, a_e^* = \arg\min \left[\mathcal{T}_o \left(a_e^{(i)} \right) + \alpha \mathcal{T}_d \left(a_e^{(i)} \right) \right] \quad \alpha \in (0, 1), \qquad (24)
$$

where $Path^*, a_e'^*$ are the final path and edge node of vehicle travel in the grid area. α is a weight parameter because $\mathcal{T}_o$ is accurately calculated using prediction information and $\mathcal{T}_d$ is an estimated value.

In addition, for the situation that the destination is not within the gridded area, the gridded area will be re-divided, traffic prediction information will be updated and path will be calculated again when the vehicle completes the path in the current gridded area. The process of the entire path planning method is shown in Algorithm 1.

Algorithm 2: Gridded Path Planning.

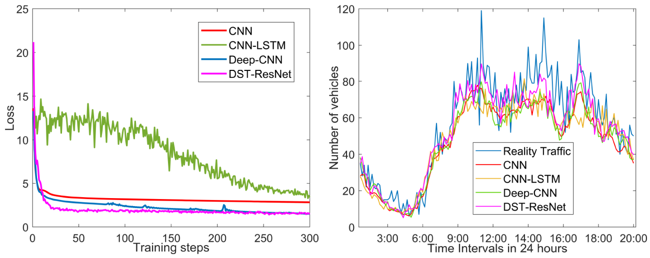
Input: Road network topology matrix R, L, Q
Prediction information V^p
Origin, destination and Departure time o, d, t_0
Searching area $\mathbb{B}$ and estimated time to d $\mathcal{T}_d$
Output: $Path^*, a_e^*$
1: Divide the grid area:
$\mathbb{A}_{grid} : \{a_i | \ |o, a_i| \leq r \quad o, a_i \in \mathbb{B}\}$
2: **if** $d \notin \mathbb{A}_{grid}$ **then**
3: Set the edge node of $\mathbb{A}_{grid}$ as $a_e^{(i)}, i = 1, 2, 3 \cdots$
4: **Initialize** $Finish = \{o\}, \mathcal{T}_o(o) = 0$
$\mathcal{T}_o(a_i) = L_{o,a_i} (R_{o,a_i} \in R)$
$\mathcal{T}_o(a_i) = \infty (a_i \neq o\ and\ R_{o,a_i} \notin R)$
5: **while** $\mathbb{A}_{grid} \neq Finish$ **do**
6: **for each** $a_i = arg\ min\ \mathcal{T}_o(a_i), a_i \notin Close$
7: **for each** $a_i^{(j)} \in neighbors\ of\ a_i$
8: Upgrade $\mathcal{T}_o(a_i^{(j)})$ by formula (20)
9: **end for**
10: $Finish = Finish \cup a_i$
11: **end for**
12: Receive $Path^*, a_e^*$ according to (24)
13: After vehicle finished the path within grid area:
14: Upgrade o and go back to step 1.
15: **else**
16: Calculate $\mathcal{T}_o(d)$ similar to step 4-10.
17: Receive $Path^*$ according to $\mathcal{T}_o(d)$.
18: **end if**

VI. Simulation

In this section, we illustrate the simulation of the fine-grained traffic prediction and the proposed gridded path planning method respectively. Specifically, we conduct the simulation of traffic prediction using 251.5 million pieces of actual data from 12587 vehicles whthin 8 days in a certain area of Beijing city with size of 31.4 km $*$ 32.1 km. Besides, the gridded path planning is simulated based on the digital map data of this area downloaded from OpenStreetMap.

A. Traffic Prediction

In the following, the simulation of fine-grained regional traffic prediction is performed. First of all, we convert the traffic data mentioned above into tensors accoding to Definitions 1 and 2. Specifically, taking 10 minutes as a time interval, we process the actual data of each time interval into a tensor $X \in \mathbb{R}^{4*100*100}$ according to vehicles' latitude, longitude and the driving direction of the data. Then, tensors data of the 7 consecutive days and data of another day will respectively be used for the training and testing of the neural networks in traffic prediction. In detail, we use the historical data of 3 recently adjacent time intervals to predict the traffic for the next time interval, which means that the input of the neural network is the historical traffic information of the last three time intervals, and the final output is the traffic information of the next time interval.



(a) (b)

Fig. 6. Comparison of training process and predicted results among different deep learning models.

TABLE I
SUMMARY OF MAIN NOTATIONS

Notation	Description
Q_i	intersection i
R_{ij}	road segment from Q_i to Q_j
L_{ij}	the length of R_{ij}
$\boldsymbol{p}_i$	information of vehicle i
t_s	time interval of traffic prediction
s_i	driving direction interval i
$x_{m,n}^{s_i,t}$	traffic flow of region (m, n) with direction s_i at time t
X_t	square maps of traffic flow at time t
v_{max}	maximum speed of roads
λ_{max}	maximum capacity of roads
o	source intersection
d	destination intersection
$\mathbb{B}$	intersection set for path searching
$\mathbb{A}_{grid}$	intersection set of grid area for path calculation
$V_{ij}^{p}(t)$	travel speed for vehicles from i to j at time t according to traffic prediction
$V_{ij}^{h}(t)$	travel speed for vehicles from i to j at time t according to historical traffic information
$\mathcal{T}_o$	accurate time from origin to the edge nodes
$\mathcal{T}_d$	estimation time from edge nodes to destination

Fig. 6 displays the comparisons of DST-ResNet and other deep learning models [18], [28], [38], in which Fig. 6(a) shows the convergence curves of training process of different deep learning models and Fig. 6(b) shows the prediction results of a certain road for 24 hours under different deep learning models.

Fig. 6(a) shows that the training of DST-ResNet converges faster, and the training loss of traffic prediction is lower than other deep learning models. Although the convergence loss of deep-CNN is similar to DST-ResNet, DST-ResNet reaches the state of convergence faster, and the loss of Deep-CNN exists fluctuations after its convergence. Also, as mentioned above, we use the data of another day that do not used in the training process to verify the prediction results, of which the regional test loss of different deep learning models are shown in Table II. It is more valid that the performance of DST-ResNet is better than others.

Fig. 6(b) illustrates the road-level testing results of traffic prediction based on different deep learning models and the real situation. It can be seen that during the 24 hours in one day, the traffic flow of a certain road varies a lot over time. For example, the traffic flow at around 3:00 AM is quite light and begin to rise, the traffic at 6:00 PM is heavier and fast-changing, and it is heavy and close to the peak at 9:00 AM. Besides, from the perspective of prediction results, the prediction results using DST-ResNet is closer to the actual situation. Table II shows the loss between the prediction results of a certain road and the actual situation, in which the loss of the prediction by DST-ResNet is the smallest.

TABLE II
COMPARISON AMONG DIFFERENT DEEP LEARNING MODELS

Deep Learning Models	Regional Loss	Loss of a single road
CNN [18]	3.6031	10.534
CNN-LSTM [38]	2.7015	10.516
Deep-CNN [28]	2.3939	10.505
DST-ResNet	1.8591	8.8175

B. Path Planning

In this subsection, the simulation result of the proposed gridded path planning method is illustrated. First of all, we download the digital map data from OpenStreetMap, process it into the form of topology matrix, including the nodes in the road network, the connection relationship between the nodes, and record the length, angle and direction of the roads. Secondly, we divide the data used in the traffic prediction above into three parts according to Definitions 3, 4 and 5. For example, if a vehicle is going to travel from an origin o to a destination d at time t, the data of the nearest time interval before time t is real-time information, the previous information is historical information, and the prediction information is the information of the next time interval predicted based on the previous two kinds of information.

At last, we randomly select multiple pairs of nodes as the origin and destination of vehicle's path planning and compare the differences for vehicles traveling from these origins to destinations at different departure times using different path planning methods, including gridded path planning (GPP) method with prediction information, the shortest path algorithm Dijkstra with real-time traffic information and Dijkstra without any traffic information. The specific content of the comparison is as follows:

- *Consuming time:* Consuming time refers to the time it takes for the vehicle to reach its destination from the origin. In this paper, the consuming time is calculated according to the actual traffic data of the entire process.
- *Traveling distance:* Traveling distance is the distance of the path traveled by the vehicle from origin to destination, which is the sum of the lengths of the roads of selected path.
- *Driving speed:* In this study, driving speed is the average speed of the vehicle during the entire driving process from origin to destination, specifically the total driving distance divided by the consuming time.
- *Calculation time:* The calculation time means the time it takes for the algorithm to finish the path calculation. The high complexity of the Dijkstra algorithm in the face of large-scale path calculations has always been a problem.

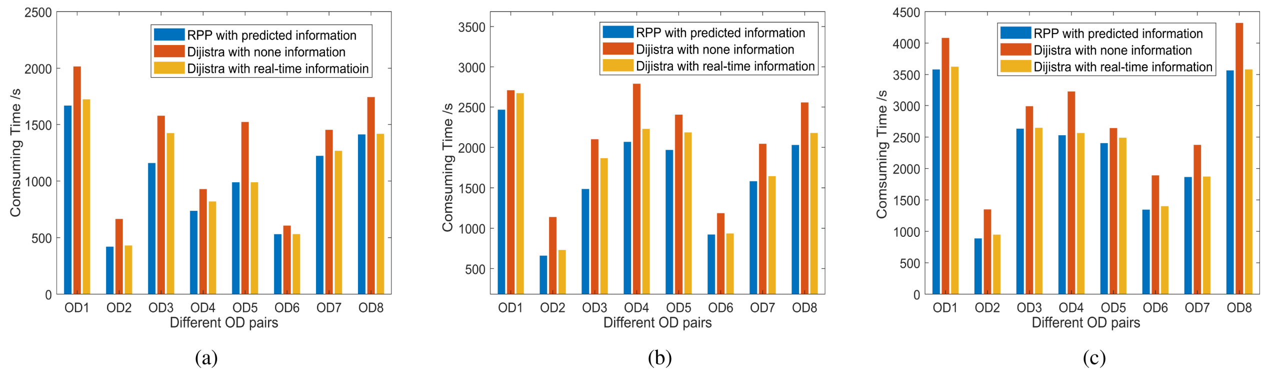


Fig. 7. Consuming time of different OD pairs at different departure times under GPP with prediction information, Dijkstra without traffic information and Dijkstra with real-time information.

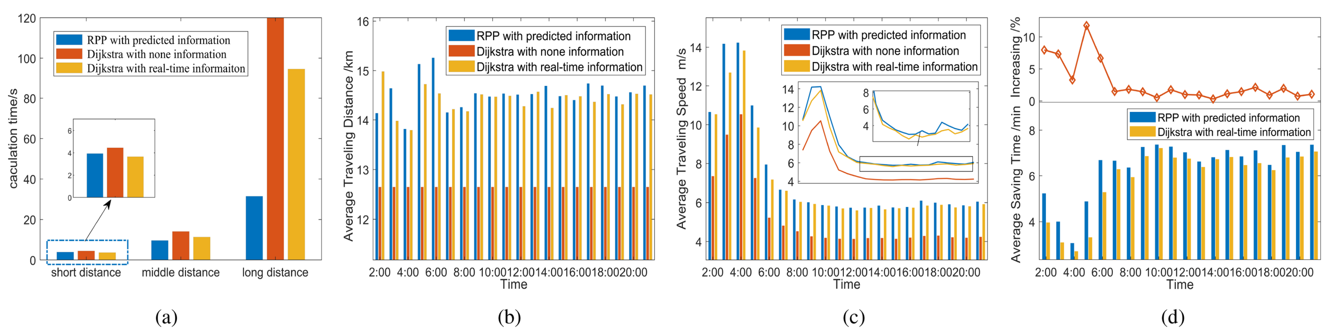


Fig. 8. Comparison of various contents under GPP with prediction information, Dijkstra without traffic information and Dijkstra with real-time information.

The specific simulation results are described in detail below, which are shown in Figs. 7 and 8.

Fig. 7(a), (b) and (c) show the consuming time for vehicles depart at 3:00, 6:00 and 9:00 AM with different pairs of origin and destination. Specifically, the simulation results illustrates that the consuming time using GPP with prediction information is less than other methods, while the average consuming time is increasing over time due to the increasing traffic flow at 3:00, 6:00 and 9:00. Besides, we can find that the performance of GPP with prediction information is better at 6:00, and has less effect at 9:00. Combined with the changes of traffic flow in the 24-hour in Fig. 6(b), this situation illustrates that GPP preforms better when the traffic change rapidly.

Fig. 8 shows the simulation results of other comparison content described above. Fig. 8(a) shows the calculation time of GPP, Dijkstra with real-time information and Dijkstra without traffic information separately under path planning demands with short, middle and long distance. Since iterative methods are used in the calculation process of all the algorithms, in the case of short and medium distances, the calculation time of these three methods do not have many differences. However, when travel distances becomes larger, the calculation time of GPP is significantly smaller than other methods, because the calculation amount of iteration increase exponentially as the number of nodes increases, and GPP divide the entire path planning process into multiple gridded path planning. Therefore, GPP has a relatively low complexity and can achieve real-time calculation easier.

Fig. 8(b) shows the average traveling distance of vehicles under different path planning methods at different departure times. It can be seen that the average traveling distance is always the smallest and consistent at different departure times under Dijkstra without traffic information, while the traveling distance under GPP and Dijkstra with real-time information is larger, which means that traffic information can help to find the path with longer distance but lower consuming time.

From Fig. 8(c), the average travel speed by Dijkstra without traffic information is the lowest, and becomes higher after having real-time traffic information. Moreover, although real-time information has greatly help to find a path with lower consuming time on the basis of Dijkstra algorithm, the average travel speed using GPP with prediction information is even higher.

Finally, Fig. 8(d) shows the time saved by Dijkstra with real-time information and GPP with prediction based on Dijkstra without traffic information and shows percentage of improvement using GPP with prediction information based on Dijkstra with real-time information at different times. It can be discovered that the time saved by GPP is always larger and it reaches the peak at around 6 AM.

VII. Conclusion

In this paper, we have proposed a DL based fine-grained urban traffic prediction framework and a gridded path planning method based on prediction information for connected vehicular networks. In the proposed method, vehicles' path can be periodically planned in each grid area according to traffic prediction information. We have performed the simulation on our proposed method using real traffic data and digital map of Beijing city. The results have confirmed that the fine-grained traffic prediction can predict the traffic flow of each road in urban road network better than other deep learning method. At the same time, the gridded path planning has reduced the average consuming time of vehicles and improved the average travel speed of vehicles compared to the Dijkstra algorithm without traffic information and Dijkstra with real-time traffic information. For the future work, we will conduct extensive AI enabled traffic predictions and apply traffic prediction to provide personalized services for connected vehicular networks.

References

[1] P. Mittal, Y. Singh, and Y. Sharma, "Analysis of dynamic road traffic congestion control (DRTCC) techniques," *Amer. J. Eng. Res.*, vol. 4, no. 10, pp. 40–47, 2015.

[2] B. Qian, H. Zhou, F. Lyu, J. Li, T. Ma, and F. Hou, "Toward collision-free and efficient coordination for automated vehicles at unsignalized intersection," *IEEE Internet Things J.*, vol. 6, no. 6, pp. 10408–10420, Dec. 2019.

[3] *T. Transp, Inst.(2007). Texas transportation institute: Urban mobility information annual urban mobility report*, College station, TX, USA.

[4] M. Saberi et al., "A simple contagion process describes spreading of traffic jams in urban networks," *Nature Commun.*, vol. 11, 2020, Art. no. 1616.

[5] M. Zhu et al., "Public vehicles for future urban transportation," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 12, pp. 3344–3353, Dec. 2016.

[6] F. Tang, Y. Kawamoto, N. Kato, and J. Liu, "Future intelligent and secure vehicular network toward 6G: Machine-learning approaches," *Proc. IEEE*, vol. 108, no. 2, pp. 292–307, Feb. 2020.

[7] Y. Wang, J. Zheng, and N. Mitton, "Delivery delay analysis for roadside unit deployment in vehicular ad hoc networks with intermittent connectivity," *IEEE Trans. Veh. Technol.*, vol. 65, no. 10, pp. 8591–8602, Oct. 2016.

[8] H. Zhou, W. Xu, J. Chen, and W. Wang, "Evolutionary V2X technologies toward the Internet of Vehicles: Challenges and opportunities," *Proc. IEEE*, vol. 108, no. 2, pp. 308–323, Feb. 2020.

[9] S. Joerer et al., "Enabling situation awareness at intersections for IVC congestion control mechanisms," *IEEE Trans. Mobile Comput.*, vol. 15, no. 7, pp. 1674–1685, Jul. 2016.

[10] L. Gao, Y. Xu, and X. Wang, "MAP: Multiauctioneer progressive auction for dynamic spectrum access," *IEEE Trans. Mobile Comput.*, vol. 10, no. 8, pp. 1144–1161, Aug. 2011.

[11] B. Qian et al., "Leveraging dynamic Stackelberg pricing game for multi-mode spectrum sharing in 5G-VANET," *IEEE Trans. Veh. Technol.*, vol. 69, no. 6, pp. 6374–6387, Jun. 2020.

[12] M. Chen, X. Yu, and Y. Liu, "Mining moving patterns for predicting next location," *Inf. Syst.*, vol. 54, pp. 156–168, 2015.

[13] H. Zhou et al., "WhiteFi infostation: Engineering vehicular media streaming with geolocation database," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 8, pp. 2260–2274, Aug. 2016.

[14] S. Verma, Y. Kawamoto, Z. M. Fadlullah, H. Nishiyama, and N. Kato, "A survey on network methodologies for real-time analytics of massive IoT data and open research issues," *IEEE Commun. Surv. Tut.*, vol. 19, no. 3, pp. 1457–1477, Jul.–Sep. 2017.

[15] F. G. Habtemichael and M. Cetin, "Short-term traffic flow rate forecasting based on identifying similar traffic patterns," *Transp. Res. Part C Emerg. Technol.*, vol. 66, no. May, pp. 61–78, 2016.

[16] M. Chen, X. Yu, and Y. Liu, "PCNN: Deep convolutional networks for short-term traffic congestion prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 11, pp. 3550–3559, Nov. 2018.

[17] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[18] J. Li et al., "An end-to-end load balancer based on deep learning for vehicular network traffic control," *IEEE Internet Things J.*, vol. 6, no. 1, pp. 953–966, Feb. 2019.

[19] Y. Xu, Q.-J. Kong, R. Klette, and Y. Liu, "Accurate and interpretable Bayesian MARS for traffic flow prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 6, pp. 2457–2469, Dec. 2014.

[20] E. W. Dijkstra et al., "A note on two problems in connexion with graphs," *Numerische Mathematik*, vol. 1, no. 1, pp. 269–271, 1959.

[21] P. E. Hart, N. J. Nilsson, and B. Raphael, "A formal basis for the heuristic determination of minimum cost paths," *IEEE Trans. Syst. Sci. Cybern.*, vol. 4, no. 2, pp. 100–107, Jul. 1968.

[22] R. Arnott, A. De Palma, and R. Lindsey, "Does providing information to drivers reduce traffic congestion?," *Transp. Res. Part A: Gen.*, vol. 25, no. 5, pp. 309–318, 1991.

[23] L. Lv, M. Chen, Y. Liu, and X. Yu, "A plane moving average algorithm for short-term traffic flow prediction," in *Proc. Pacific-Asia Conf. Knowl. Discov. Data Mining*, 2015, pp. 357–369.

[24] M. Yildirimoglu and N. Geroliminis, "Experienced travel time prediction for congested freeways," *Transp. Res. Part B*, vol. 53, pp. 45–63, Jul. 2013.

[25] E. Ko, J. Ahn, and E. Y. Kim, "3D Markov process for traffic flow prediction in real-time," *Sensors*, vol. 16, no. 2, 2016, Art. no. 147.

[26] E. Bolshinsky and R. Friedman, "Traffic flow forecast survey," Comput. Sci. Dept., Technion, Haifa, Israel, Tech. Rep. CS-2012-06, 2012.

[27] A. Abadi, T. Rajabioun, and P. A. Ioannou, "Traffic flow prediction for road transportation networks with limited traffic data," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 653–662, Apr. 2015.

[28] X. Ma, Z. Dai, Z. He, J. Ma, Y. Wang, and Y. Wang, "Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction," *Sensors*, vol. 17, no. 4, 2017, Art. no. 818.

[29] J. Zhang, Y. Zheng, and D. Qi, "Deep spatio-temporal residual networks for citywide crowd flows prediction," in *Proc. AAAI Conf. Artif. Intell.*, 2017, pp. 1655–1661.

[30] S. Guo, Y. Lin, S. Li, Z. Chen, and H. Wan, "Deep spatial–temporal 3D convolutional neural networks for traffic data forecasting," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 10, pp. 3913–3926, Oct. 2019.

[31] R. Silva, S. M. Kang, and E. M. Airoldi, "Predicting traffic volumes and estimating the effects of shocks in massive transportation systems," *Proc. Nat. Acad. Sci. USA*, vol. 112, no. 18, pp. 5643–5648, 2015.

[32] R. Rajagopalan, K. G. Mehrotra, C. K. Mohan, and P. K. Varshney, "Hierarchical path computation approach for large graphs," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 44, no. 2, pp. 427–440, Apr. 2008.

[33] M. K. Mainali, S. Mabu, S. Yu, S. Eto, and K. Hirasawa, "Dynamic optimal route search algorithm for car navigation systems with preferences by dynamic programming," *IEEJ Trans. Elect. Electron. Eng.*, vol. 6, no. 1, pp. 14–22, 2011.

[34] M. Wang, H. Shan, R. Lu, R. Zhang, X. Shen, and F. Bai, "Real-time path planning based on hybrid-VANET-enhanced transportation system," *IEEE Trans. Veh. Technol.*, vol. 64, no. 5, pp. 1664–1678, May 2015.

[35] C. Guo, D. Li, G. Zhang, and M. Zhai, "Real-time path planning in urban area via VANET-assisted traffic information sharing," *IEEE Trans. Veh. Technol.*, vol. 67, no. 7, pp. 5635–5649, Jul. 2018.

[36] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[37] H. Zhou, N. Cheng, Q. Yu, X. S. Shen, D. Shan, and F. Bai, "Toward multi-radio vehicular data piping for dynamic DSRC/TVWS spectrum sharing," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 10, pp. 2575–2588, Oct. 2016.

[38] O. Vinyals, A. Toshev, S. Bengio, and D. Erhan, "Show and tell: A neural image caption generator," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3156–3164.

Hanlin Wu (Student Member, IEEE) received the B.S. degree in electronic information science and technology in 2019 from NanJing University, Nanjing, China, where he is currently working toward the M.S. degree with the School of Electronic Science and Engineering. His research interests include intelligent transportation systems and vehicular networks, satellite networks, and machine learning applications for wireless communication.

Haibo Zhou (Senior Member, IEEE) received the Ph.D. degree in information and communication engineering from Shanghai Jiao Tong University, Shanghai, China, in 2014. From 2014 to 2017, he was a Postdoctoral Fellow with the Broadband Communications Research Group, Department of Electrical and Computer Engineering, University of Waterloo, Waterloo,ON, Canada. He is currently an Associate Professor with the School of Electronic Science and Engineering, Nanjing University, Nanjing, China. His research interests include resource management and protocol design in vehicular ad hoc networks, cognitive networks, and space-air-ground integrated networks. He was the recipient of the 2019 IEEE ComSoc Asia-Pacific Outstanding Young Researcher Award. He was an Invited Track Co-Chair for ICCC'2019, VTC-Fall'2020 and a TPC Member of many IEEE conferences, including GLOBECOM, ICC, and VTC. He was an Associate Editor for the IEEE Comsoc Technically Co-Sponsored the *Journal of Communications and Information Networks* from April 2017 to March 2019, and a Guest Editor of the *IEEE Communications Magazine* in 2016, the Hindawi International *Journal of Distributed Sensor Networks* in 2017, and *IET Communications* in 2017. He is currently an Associate Editor for the IEEE INTERNET OF THINGS JOURNAL, the *IEEE Network Magazine*, and the IEEE WIRELESS COMMUNICATIONS LETTER.

Jiwei Zhao (Student Member, IEEE) received the M.S. degree in information and communication system from Xidian University, Xi'an, China, in 2018. He is currently working toward the Ph.D. degree with the School of Electronic Science and Engineering, Nanjing University, Nanjing, China. He won the first prize in the 2016 CCF (China Computer Federation) China Big Data & Cloud Computing Intelligence Contest. His research interests include fully-decoupled RAN architecture, coordinated multi-point, and machine learning applications for wireless communication.

Yunting Xu (Student Member, IEEE) received the B.S. degree in communication engineering in 2017 from Nanjing University, Nanjing, China, where he is currently working toward the Ph.D. degree with the School of Electronic Science and Engineering. His research interests include dynamic resource management and networking optimization in the field of emerging wireless networks.

Bo Qian (Student Member, IEEE) received the B.S. and M.S. degrees in statistics from Sichuan University, Chengdu, China, in 2015 and 2018, respectively. He is currently working toward the Ph.D. degree with the School of Electronic Science and Engineering, Nanjing University, Nanjing, China. His research interests include intelligent transportation systems and vehicular networks, wireless resource management, blockchain, convex optimization theory, and game theory. He was the recipient of the Best Paper Award from IEEE VTC2020-Fall.

Xuemin (Sherman) Shen (Fellow, IEEE) received the Ph.D. degree in electrical engineering from Rutgers University, New Brunswick, NJ, USA, in 1990. He is currently a University Professor with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. His research interests include network resource management, wireless network security, Internet of Things, 5G and beyond, and vehicular ad hoc, and sensor networks. He is a registered Professional Engineer of Ontario, Canada, an Engineering Institute of Canada Fellow, a Canadian Academy of Engineering Fellow, a Royal Society of Canada Fellow, and a Distinguished Lecturer of the IEEE Vehicular Technology Society and Communications Society.

Dr. Shen was the recipient of the R.A. Fessenden Award in 2019 from IEEE, Canada, Award of Merit from the Federation of Chinese Canadian Professionals (Ontario) presents in 2019, James Evans Avant Garde Award in 2018 from the IEEE Vehicular Technology Society, Joseph LoCicero Award in 2015 and Education Award in 2017 from the IEEE Communications Society, Technical Recognition Award from Wireless Communications Technical Committee (2019) and AHSN Technical Committee (2013), Excellent Graduate Supervision Award in 2006 from the University of Waterloo, and Premier's Research Excellence Award (PREA) in 2003 from the Province of Ontario, Canada. He was the Technical Program Committee Chair/Co-Chair of the IEEE Globecom'16, IEEE Infocom'14, IEEE VTC'10 Fall, IEEE Globecom'07, the Symposia Chair of the IEEE ICC'10, and the Chair of the IEEE Communications Society Technical Committee on Wireless Communications. He is the elected IEEE Communications Society Vice President for Technical & Educational Activities, Vice President for Publications, Member-at-Large on the Board of Governors, Chair of the Distinguished Lecturer Selection Committee, Member of the IEEE Fellow Selection Committee. He is the Editor-in-Chief of IEEE INTERNET OF THINGS JOURNAL, IEEE NETWORK, *IET Communications*, and *Peer-to-Peer Networking and Applications*.